

# Guiding learners to near native fluency in English through an adaptive programme of activities which includes phoneme and prosody analysis

Alistair Lawson<sup>1</sup>, Ann Attridge<sup>2</sup>, and Paul Lapok<sup>3</sup>

**Abstract.** Many students of English language find pronunciation difficult to master. This work in progress paper discusses an incremental and iterative approach towards developing requirements for software applications to assist learners with the perception and production of English pronunciation in terms of phonemes and prosody. It was found that prompts for eliciting target pronunciation should include a visual indication of the meaning of the target word or phrase in addition to the sound, and that the learners should be led through a hierarchy of words. This should start with phonemes of simple (one syllable) words, and adaptively build up to prosody of two syllable words, then increasing the syllables in the target words as the learner improves. A simple representation of prosody was developed and found to be intuitive by students for comparing their pronunciation to that of a native speaker. Students considered that an analysis time of within one second for phonemes and prosody was considered “real time”, and requested integration with social media for both enabling competition and celebration of achievement.

**Keywords:** MALL, ESOL, pronunciation, prosody, feedback.

---

1. Institute for Informatics and Digital Innovation, Edinburgh Napier University; a.lawson@napier.ac.uk.

2. K2L Ltd United Kingdom.

3. Institute for Informatics and Digital Innovation, Edinburgh Napier University.

**How to cite this article:** Lawson, A., Attridge, A., & Lapok, P. (2014). Guiding learners to near native fluency in English through an adaptive programme of activities which includes phoneme and prosody analysis. In S. Jager, L. Bradley, E. J. Meima, & S. Thoučsny (Eds), *CALL Design: Principles and Practice; Proceedings of the 2014 EUROCALL Conference, Groningen, The Netherlands* (pp. 191-195). Dublin: [Research-publishing.net](http://Research-publishing.net). doi:10.14705/rpnet.2014.000216

## 1. Introduction

This project aims to address pronunciation problems of English language learners. Many students of English at all levels find pronunciation difficult to learn. Though important for comprehension and fluency, pronunciation is seen by many as being given the least attention in language learning (e.g. Gilakjani, Ahmadi, & Ahmadi, 2011). Achievement of near native fluency involves the ability to reproduce English prosody in terms of pitch, intensity, and duration, in addition to basic phonemic competence. However, current pronunciation software tools mainly address phonemic difficulties and give little or no analytical feedback, or too much feedback, such as complex graphs of speech waveforms and spectrograms, but in a way that pays little attention to problems with prosody. This work in progress paper reports on the preliminary results of a project entitled Protalk, which includes phonemic diagnosis but also takes the learner forward by analysing and giving usable feedback on prosody problems. The project, with a view to developing mobile apps, is being carried out in an iterative and incremental software development approach (e.g. Demetris, Farnum, Markel, & Rosenhan, 2012) with a focus on user experience design and evaluation, market research, and with a multidisciplinary team of language teachers, software engineers, games developers, and marketing professionals.

## 2. Method

The first stage of user evaluation focussed on intelligibility of prompts used to illicit the pronunciation of target words or phrases by the student. This focussed solely on prosody and investigated ten subjects (five intermediate to advanced, and five beginners), whose pronunciation of a set of predetermined words, phrases and sentences were benchmarked against a set of native speaker recordings in order to establish the following:

- The extent to which the learner benefitted from audio only or audio plus text as a prompt to pronouncing the words, in order to establish the feasibility of training ear and vocal apparatus without text-based prompts.
- To what extent the learner's pronunciation was affected by their understanding of the words they were pronouncing in order to establish the most effective learning methodology when using de-contextualised speech segments.

The second stage of user evaluation involved investigating how to present the potentially complex feedback in such a way that learners understand, engage

and are motivated to improve. This used mock-ups of a mobile application, and the paths that a student would take through the learning experience. Ten further subjects took part in this analysis. A set of simple visual symbols was devised which represent pitch, intensity, duration, and give learners instant visual feedback on all three parameters in a simple, clear manner that avoids information overload, which can occur with existing methods of displaying waveforms and spectrograms.

In parallel with these two evaluations, a cloud-based web service API (Application Programming Interface) was developed to allow analysis of speech recorded by the learners to be analysed for phonemic accuracy, and for pitch, intensity, and syllable duration. The development of this web service API was carried out in an iterative and incremental manner taking account of the findings of the user-centred evaluations, two of which are reported here. In addition, a market validation was conducted and the findings incorporated into the design requirements.

### **3. Results**

The first evaluation identified the use of audio as the prime focus, but some learners expressed a preference for the written word. As a result, the option of accessing the text after the audio was included in the design of the mobile application. The second finding indicated that for all students it was beneficial for pronunciation to understand the meaning of the target speech segment.

Additional findings from the first evaluation (with the beginners group) firstly made it obvious that it was necessary to devise an adaptive system to determine the individual phonemes that learners were struggling with in one or two syllable words before moving on to analysing prosody alongside phonemic diagnosis in more complex words and phrases. The market validation report also confirmed the need to include phonemic analysis as a starting point. There was a need to (1) provide contextualisation and aid understanding through illustration (using images) and access to dictionary support as required, and (2) cater for different levels of students by creating sets of words appropriate to their lexical knowledge.

The findings of the second evaluation (for the intermediate level group) were that pronunciation was the most challenging part of language learning (as compared to learning vocabulary and grammatical structure). This evaluation confirmed that most students wanted to master the phonemes first before moving on to improving their prosodic ability. The division of words into syllable groups, and progressing through levels from two, to three, to four syllable words was thought to be helpful in building up prosodic competence. Feedback was also received on the look and

feel of the application, and in particular the representation of prosody, but also included requesting the use of animations in addition to static images to help give the meaning of the target words, and make the app more attractive and engaging. A simple representation of prosody was developed and found to be intuitive by students for comparing their pronunciation to that of a native speaker. Integration with social media was requested by students for competing against other learners or celebrating achievement, as was the ability to track performance over time. A response time for analysis of the speed recordings of within one second was considered to be “real time” by users.

#### **4. Discussion and future work**

The main areas of challenge relating to the development of this kind of mobile app included:

- the design of intelligible prompts for eliciting the target pronunciation;
- the design of the appropriate learning paths through a hierarchy of target pronunciations;
- the quantity of feedback required by the learner;
- the quality of feedback required by the learner;
- the speed of feedback required by the learner.

The current project focusses on providing feedback in relation to phonemic pronunciation and three components of prosody: loudness, pitch and duration of syllable for individual words. Future work will involve investigating how to include phrases and sentences. The accuracy of the analysis results provided by the Web service API requires evaluation and benchmarking against a database of words and phrases that have been accurately tagged for phonemic and prosodic features. The robustness, scalability and security of the webservice will also need to be evaluated.

Applications of this technology include mobile apps (e.g. for single words), call centre training (e.g. customised scripts), and a children’s adventure game (e.g. to engage children in mastering English pronunciation).

#### **5. Conclusions**

The user centred approach was useful in determining requirements for assisting learners with the perception and production of English pronunciation. Prompts for eliciting target pronunciation should include a visual indication (such as pictures or

animations) of the meaning of the target word or phrases in addition to the sounds. Learners should be led through a hierarchy of words, starting with phonemes of one syllable words as the target, and adaptively building up to that prosody of two syllable words, then increasing to three syllable words, and so on, as the learner improves. A simple representation of prosody was developed and found to be intuitive by students for comparing their pronunciation to that of a native speaker. An analysis time of within one second for phonemes and prosody was considered “real time” by students. Integration with social media for both enabling competition and celebration of achievement was requested by participants.

**Acknowledgements.** This work results from the Protalk project funded by the UK Technology Strategy Board.

## References

- Demetris, C., Farnum, C., Markel, J., & Rosenhan, S. (2012, June 4). Developing UX agility: Letting go of perfection. *UX Matters*. Retrieved from <http://www.uxmatters.com/mt/archives/2012/06/developing-ux-agility-letting-go-of-perfection.php>
- Gilakjani, A., Ahmadi, S., & Ahmadi, M. (2011). Why is pronunciation so difficult to learn? *English Language Teaching*, 4(3), 74.